

Der Einfluss von Stimmung auf die Lehrevaluation

Jörg Zumbach und Joachim Funke

Die Evaluation der Lehre gehört zu einem Standardinstrument der Qualitätssicherung an deutschen und österreichischen Universitäten. Entsprechend hat sich auch eine Vielzahl an Verfahren etabliert, welche zum Teil recht heterogene Dimensionen erfassen und auch unterschiedliche methodische Zugänge aufweisen. Gerade etwa bei der Methodik gibt es unterschiedliche Strategien, um systematische Störeinflüsse bei der Beurteilung von Lehrveranstaltungen zu vermeiden. Eine experimentell recht gut beforschte Variable, welche sich auf die Urteilsbildung auswirkt, ist die Stimmung von Urteilenden. In einem Quasi-Experiment haben wir untersucht, wie sich unterschiedliche Stimmung auf die Beurteilung von Vorlesungen auswirkt. In mehreren Vorlesungen an den Universitäten Salzburg und Heidelberg wurde mithilfe der Facial-Feedback-Hypothese dieser Einfluss untersucht. Während der Beurteilung der Lehrveranstaltungen wurde bei einer Hälfte der Studierenden mittels eines Stiftes zwischen den Zähnen eine positive Stimmung induziert, bei der zweiten Hälfte mittels eines Stiftes zwischen den Lippen eine eher neutrale/negative Stimmung. Die Ergebnisse zeigen, dass eine positive Stimmung zu einer signifikant besseren Bewertung führt als eine neutrale/negative Stimmung. Wenngleich der Effekt zwar nur eine Viertelpunkte groß ist, belegt dies dennoch die systematische Beeinflussbarkeit von Lehrevaluationen.

Hintergrund und Fragestellung

Die Evaluation der Lehre gehört mittlerweile zu einem Standardinstrument der Qualitätssicherung an deutschen und österreichischen Universitäten. Entsprechend hat sich auch eine Vielzahl an Verfahren etabliert, welche zum Teil recht heterogene Dimensionen erfassen und auch unterschiedliche methodische Zugänge aufweisen. Gerade etwa bei der Methodik gibt es unterschiedliche Strategien, um systematische Störeinflüsse bei der Beurteilung von Lehrveranstaltungen zu vermeiden. So werden spezifische Auswertungsstrategien eingesetzt oder auch in der Semestermitte evaluiert, um potenzielle Abbrecher zu erfassen (z.B. Zumbach, Spinath & Schahn, 2007). Systematischen Verzerrungen werden auch hinsichtlich der „Beliebtheit“ eines Fachs oder einer Disziplin diskutiert. Eine experimentell recht gut beforschte Variable, welche sich auf Urteilsbildung auswirkt, ist die Stimmung von Urteilenden

(z.B. Schwarz, Bless & Bohner, 1991). Hierzu liegt noch wenig Evidenz aus dem Bereich der Qualitätssicherung vor.

Stimmung und Informationsverarbeitung

Im vorliegenden Experiment sollte der Frage nachgegangen werden, ob und wie sich unterschiedliche Stimmungen auf die Evaluation von Lehrveranstaltungen auswirken. Anzunehmen ist, dass bei positiver Stimmung eher eine oberflächliche Informationsverarbeitung stattfindet. Diese sollte insgesamt zu einer positiveren Bewertung als eine eher negative Stimmung führen.

Hintergrund dieser Hypothese ist die von Fiedler (2001) vorgeschlagene Differenzierung von ASSIMILATIVER und AKKOMMODATIVER Verarbeitung, die unterschiedlich von positiven oder negativen Emotionen beeinflusst werden. Assimilative Verarbeitung bezieht sich auf Prozesse der Transformation und aktiven Elaboration von Stimulusinformationen, während akkommodative Verarbeitung eine möglichst exakte Präservation der Informationen zum Ziel hat, wobei die Vermeidung von Fehlern wichtiger ist als das Herstellen neuer Verbindungen und Interpretationen. Fiedler zu Folge entsprechen diese beiden Verarbeitungsstile den Lerhsets für appetitive und aversive Situationen: Appetitive Situationen werden am besten durch Erforschung und kreatives Verhalten genutzt, während in aversiven Situationen vorsichtiges Verhalten, Informationssuche und Fehlervermeidung sinnvoll sind. Da diese Situationen typischerweise von positivem oder negativem Affekt begleitet werden, aktiviert der entsprechende Affekt nun wiederum eine dem ursprünglichen Lernset ähnliche Verarbeitung: Positiver Affekt aktiviert assimilative Verarbeitung, während negativer die akkommodative Verarbeitung anregt.

Daraus folgt, dass positive Stimmung in kreativen, generativen Verarbeitungsprozessen unterstützend wirkt; allerdings kann die informationsbewahrende, vorsichtige Strategie, welche durch negative Stimmung angeregt wird, je nach Aufgabenanforderung auch Vorteile haben, was der gängigen Auffassung widerspricht, dass negative Emotionen grundsätzlich hemmend wirken. In Bezug auf Evaluationskontexte wird erwartet, dass durch negative Stimmung der Blick auf kritische Details der Lehrveranstaltung gerichtet wird, während positive Stimmung eher den globalen, zumeist guten Eindruck in pauschaler Form fördert.

Der Einfluss von Stimmung auf Lehrevaluationen

Um nun den systematischen Einfluss von Stimmung auf die Beurteilung von Lehrveranstaltungen zu untersuchen, wurde in verschiedenen Vorlesungen mittels der *Facial-Feedback*-Hypothese eine positive bzw. negative Stimmung induziert. Einen Einfluss von induzierter Stimmung durch gezielte Muskelkontraktionen auf Urteilsprozesse wird dabei angenommen und baut auf Befunde in diesem Bereich auf (vgl. Bodenhausen, Kramer & Süßner, 1994; Strack & Neumann, 2000). Während der Stimmungsinduktion hier wurde dabei seitens der Probanden ein Fragebogen zur Lehrevaluation bearbeitet. Zu dieser Evaluation wurde der „Fragebogen zur Evaluation von Vorlesungen“ des Psychologischen Institutes der Universität Heidelberg verwendet. Dabei handelt es sich um eine Weiterentwicklung der Skala nach Zumbach et al. (2007). Insgesamt 39 Items erfassen dabei sieben unterschiedliche Dimensionen: (1) Wissenserwerb, (2) Bewertung des Dozierenden, (3) Rahmenbedingungen, (4) Anforderung, (5) Inhalte, (6) Teilnehmende, (7) Gesamtbewertung. Die internen Konsistenzen ergaben dabei gute bis mäßige Werte (vgl. Tabelle 1; ein Item wurde bei der Gesamtbewertung zur Steigerung der internen Konsistenz ausgeschlossen).

Zur Induktion unterschiedlicher Stimmungslagen wurde auf die *Facial-Feedback*-Hypothese zurückgegriffen (vgl. Strack, Martin & Stepper, 1988). Danach reicht bereits die Veränderung bestimmter Gesichtsmuskeln, um einen entsprechen positiven oder negativen Affekt zu erzeugen. Eine Kontraktion der Lippen aktiviert den *orbicularis oris* und stimuliert Missbilligung, während das Beißen auf die Zähne den *zygomaticus major* anspannt und damit ein Lächeln produziert. Im Sinne eines *Embodied-Cognition*-Ansatzes ist die damit jeweils verbundene Mimik nicht nur Ausdruck einer bestimmten Stimmung, sondern auch gleichzeitig Information. Auf Basis einer bestimmten Muskelspannung und -entspannung kann diese Information wiederum weiterverarbeitet werden um etwa eine bestimmte Stimmung anzunehmen. Eine „lächelnde“ Mimik, bedingt durch den Stift zwischen Zähnen kann somit als Information dienen, dass man in einer guten Stimmung ist. Dies wird übernommen und führt damit tatsächlich zu einer besseren Stimmung. Ein Stift zwischen Lippen führt eher zu einer Mimik, die eine negative Stimmung ausdrückt. Entsprechend wird auch hier der korrespondierende emotionale Eindruck induziert. Im Experiment von Strack et al. (1988) wurde der Zweck der Stimmungsinduktion durch die unterschiedliche Stiftposition nicht von den Probanden erkannt und führte in der Tat zu den intendierten Stimmungsinduktionen. Es ist anzunehmen, dass auch in der vorliegenden Studie der eigentliche Sinn der Verwendung eines Stifts nicht erkannt wurde.

Tab. 1: Interne Konsistenzen (Cronbach's Alpha) der sieben verwendeten Skalen

| Skala | Cronbach's Alpha |
|-----------------|------------------|
| Wissenserwerb | 0,74 |
| Dozierende | 0,75 |
| Rahmen | 0,71 |
| Anforderung | 0,75 |
| Inhalte | 0,47 |
| Teilnehmende | 0,48 |
| Gesamtbewertung | 0,60 |

Per Zufallsprinzip wurden die Probanden im Hörsaal gebeten gebeten, einen Stift entweder zwischen die Lippen oder zwischen die Schneidezähne zu klemmen und dies während der Fragebogenbearbeitung beizubehalten. Abbildung 1 zeigt eine typische Haltung während der Bearbeitung des Evaluationsbogens.

Insgesamt nahmen 109 Studierende an den Universitäten Salzburg und Heidelberg im Rahmen von zwei Vorlesungen teil (86 Frauen und 23 Männer).



Abb. 5: Stimmungsinduktion bei der Fragebogenbearbeitung
Der Stift ist entweder zwischen die Lippen (negative Stimmung)
oder zwischen den Schneidezähnen (positive Stimmung)

Ergebnisse der Untersuchung

Die Ergebnisse in Abbildung 2 zeigen, dass sich beide Gruppen voneinander unterscheiden. Die Probanden mit dem Stift zwischen den Zähnen (positive Stimmungsinduktion) beurteilten die Vorlesungen insgesamt besser hinsichtlich der Gelgenheit zum Wissenserwerb, beurteilten die Dozierenden besser, die Inhalte, die anderen Teilnehmenden und gaben eine bessere Gesamtnote für die Lehrveranstaltungen ab als die Probanden mit dem Stift zwischen den Lippen (negative Stimmungsinduktion). Signifikant unterscheiden sich hier die beiden

Skalen „Teilnehmende“ und „Gesamtbewertung“ (vgl. Abb. 2). Bei den übrigen Skalen konnten keine nennenswerten Unterschiede festgestellt werden.

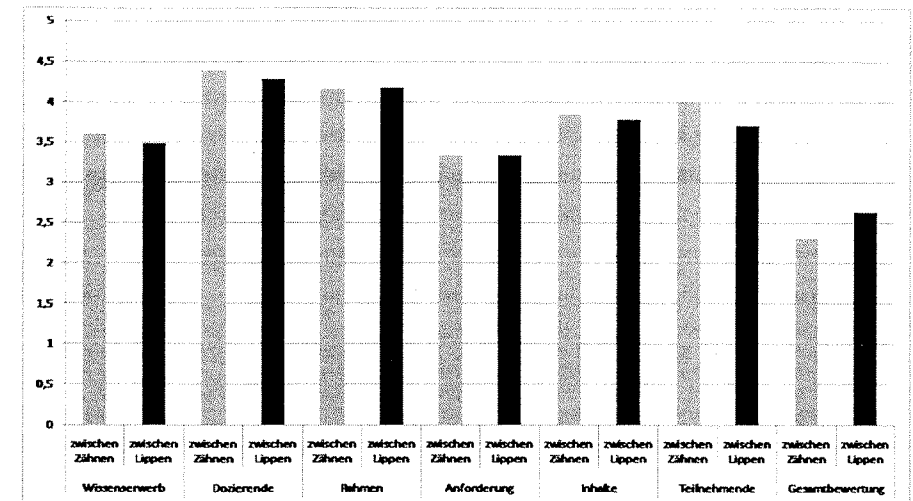


Abb. 6: Unterschiede in den sieben Skalen der Lehrveranstaltungsevaluation, jeweils zunächst der Wert für die Bedingung „zwischen Zähnen“ (helle Säulen), dann derjenige für die Bedingung „zwischen Lippen“ (dunkle Säulen)

Diskussion

Die Ergebnisse zeigen, dass Stimmung tatsächlich zu unterschiedlichen Bewertungen ein und desselben Sachverhaltes führen kann. Wenngleich hier die Stimmung experimentell durch die Facial-Feedback-Hypothese induziert wurde, können auch im Feld bei der Evaluation von Lehrveranstaltungen systematische Verzerrungen durch unterschiedliche Stimmungen verursacht werden. Im normalen Regelbetrieb ist zu erwarten dass Stimmungsschwankungen seitens der Bewertenden sich gegenseitig kompensieren, d.h. dass der Fehlerwert hier bei null liegen dürfte. Allerdings können auch systematische Störungen wie etwa Wetterlagen, Räumlichkeiten (etwa ein alter, renovierungsbedürftiger Hörsaal im Keller im Vergleich zu einem sonnendurchfluteten Neubau) oder auch Klausurankündigen etc. hier eine Rolle spielen. In dieser Studie konnten wir zeigen, dass eine positive Stimmung auch zu einer positiveren Bewertung von Lehrveranstaltungen führen kann als eine negative Stimmung. So fällt etwa die Bewertung in allen Subskalen hier besser bei positiver Stimmung aus (Stift zwischen den Zähnen).

Implikationen ergeben sich bei der Lehrveranstaltungsevaluation damit für die Ermittlung „wahrer“ Werte, also der Bewertung ohne störende Einflüsse. So könnte etwa neben der Stimmung auch etwa die Beliebtheit des Fachs erhoben und ausgerechnet werden (vgl. z.B. Reiß, Mildner, Nagler & Schweizer, 2010). Ohne solche Kontrollen können systematische Verzerrungen die Werte verfälschen. Dies ist vor allen Dingen dann problematisch, wenn anhand der Lehrveranstaltungsevaluation etwa finanzielle Ressourcen verteilt werden, wie dies an einigen Hochschulen bereits gängige Praxis ist. Dafür sind die hier aufgeführten sowie weitere potentielle Störvariablen zu erforschen, damit deren Einfluss berücksichtigt werden kann.

Literatur

- Bodenhausen, G. V., Kramer, G.P. & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, 66 (4), 621-632.
- Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. Martin & G. L. Clore (Hrsg.), *Theories of mood and cognition: A User's Guidebook* (S. 85-98). Hillsdale, NJ: Erlbaum
- Reiß, S., Mildner, D., Nagler, H. & Schweizer, K. (2010). *Teilnahmemotive an universitären Vorlesungen in Abhängigkeit vom Lehrveranstaltungsinhalt und Erfassungszeitpunkt*. Paper präsentiert bei der 8. Fachtagung Psychologiedidaktik und Evaluation, 14.-15. Mai 2010, Frankfurt a.M.
- Schwarz, N., Bless, H. & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology*, 24, 161-199.
- Strack, F., Martin, L. & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A non-obtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54, 768-777.
- Strack, F. & Neumann, R. (2000). Furrowing the Brow May Undermine Perceived Fame: The Role of Facial Feedback in Judgments of Celebrity. *Personality and Social Psychology Bulletin*, 26 (7), 762-768.
- Zumbach, J., Spinath, B. & Schahn, J. (2007). Eine Kurzsкала zur Lehrevaluation. In M. Krämer, S. Preiser & K. Brusdeylins (Hrsg.), *Psychologiedidaktik und Evaluation VI* (S. 317-126). Göttingen: V&R unipress.

Kompetenzzuwachs im Selbstbericht: Adaptation eines Instrumentes zur Evaluation von Führungskräfte Trainings

Tobias Ringeisen, Barbara Neubach, Mareike Florax und Martin Kersting

Traditionell wird der erzielte Kompetenzzuwachs durch Führungskräfte Trainings (FKTs) anhand von Fremdbereichten evaluiert. Ein geeignetes Instrument zur Selbsteinschätzung fehlt bislang, obwohl aus der Bildungsforschung ermutigende Ergebnisse für einen derartigen Evaluationsansatz vorliegen. Um diese Lücke zu schließen, wurde ein Instrument zur Evaluation der Hochschullehre – das „Berliner Evaluationsinstrument für selbst eingeschätzte studentische Kompetenzen“ (BEvaKomp; Braun, Gusy, Leidner & Hannover, 2008) für den Kontext von FKTs adaptiert. Von den sechs Subskalen wurden diejenigen vier ausgewählt, auf die FKTs im Allgemeinen abzielen (Fach-, Methoden-, Kommunikations- und Selbstkompetenz; vgl. Sonntag & Schaper, 2006). Die modifizierten Skalen wurden N = 390 Führungskräften einer Bundesverwaltung nach einem FKT vorgelegt. Konfirmatorische Faktorenanalysen und Skalenanalysen weisen auf eine befriedigende Konstruktvalidität hin. Die Interkorrelationen zwischen den einzelnen Kompetenzfacetten, Zufriedenheit und wahrgenommenem Nutzen sprechen für eine ausreichende diskriminante Validität.

Theoretischer und empirischer Hintergrund

Führungskräfte Trainings (FKTs) haben zum Ziel, den teilnehmenden Führungskräften Wissen und Fertigkeiten zum Umgang mit einer Vielzahl von Situationen aus dem Führungsalltag zu vermitteln. Abhängig vom Inhalt des jeweiligen Trainings werden verschiedene Kompetenzfacetten in unterschiedlichem Ausmaß gefördert. Diese umfassen im Allgemeinen Fach-, Methoden-, Kommunikations- und Selbstkompetenz (Sonntag & Schäfer-Rauser, 1993; Sonntag & Schaper, 2006).

In der Weiterbildungsforschung existiert eine Reihe von Ansätzen, um die Effektivität von FKTs anhand des erzielten Kompetenzzuwachses zu evaluieren (vgl. Erpenbeck & von Rosenstiel, 2007; Kaufhold, 2006). Die Mehrheit dieser Ansätze kombiniert Selbstbericht, Fremdbereicht und verhaltensbasierte Verfahren, um die Wirksamkeit eines FKT auf verschiedenen Ebenen durch die teilnehmenden Führungskräfte, ihre Mitarbeiter und/oder Vorgesetzten einschätzen zu lassen. Um Kosten und Aufwand zu reduzieren, konzentriert sich die Evaluation in der Praxis auf Selbsteinschätzungen der Trainingsteilnehmer (Bogner, 2007). Erho-